MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924
Index: Harvard, Boston, Sydney, Copernicus, Sinta 4, Scilit, Semantic, Google, etc

https://doi.org/10.58578/mikailalsys.v4i3.11980

A Novel Generative Adversarial Network Approach to Autonomous Threat Detection and Prevention in Educational Internet of Things Platforms

Onoja Emmanuel Oche[1], Suleiman Muhammad Nasir[2], Elizabeth Audu[3], Maimuna Abdullahi Ibrahim[4]

[1]Federal University of Technology Minna, Nigeria; [2,3,4]Federal Polytechnic Nasarawa, Nigeria
onoskiss@gmail.com; mnsuleimano@gmail.com

Submitted:	Revised:	Accepted:	Published:
Aug 16, 2026	Sep 13, 2026	Sep 25, 2026	Sep 30, 2026

Abstract

Educational Internet of Things (IoT) systems generate sensitive student data through real-time learning analytics, monitoring, tracking, and adaptive instruction, creating security challenges for conventional intrusion detection systems that rely on labeled attack data and exhibit high false-positive rates. This study developed and evaluated a lightweight, edge-deployable framework based on a Wasserstein generative adversarial network with gradient penalty (WGAN-GP) for autonomous threat detection and proactive prevention in educational IoT environments. The generator produced synthetic attack patterns to address extreme class imbalance, while the discriminator learned normal student-device behavior and distinguished anomalies in real time. A policy engine operating on the same edge node autonomously triggered alert escalation, device isolation, and traffic rerouting upon threat detection. The framework was implemented on an ESP32-based smart-classroom testbed using augmented CIC-IoT 2023 data and MQTT-based educational traffic traces. It achieved 98.7% detection accuracy, an average inference latency of 0.12 s, and a 94% reduction in false positives compared with conventional machine learning baselines. Its reported memory footprint was below 180 [unit to be confirmed]. These findings indicate the framework's potential to integrate real-time threat detection with automated preventive responses at the edge. The study contributes a generative adversarial network framework for autonomous defense in educational IoT environments, with practical implications for protecting sensitive student data.

Mikailalsys is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

Keywords: Generative Adversarial Networks; Educational Internet of Things; Autonomous Threat Detection; Threat Prevention; Edge Computing

Introduction

The integration of smart devices into all spheres of teaching and learning has revolutionised pedagogy (Uddin, Rahman, & Alam, 2025). Learning environments have been enhanced to a state of real-time data monitoring, evaluation, and analytics without geographical barriers. This smart learning environment generates continuous data from students' biometrics, behavioural patterns, attendance logs, and academic performance, which are often vulnerable to sophisticated attacks (Zhao, Wang, & Xu, 2024).

Although there are regulatory protections for students' data, the heterogeneous and memory-limited nature of the interconnected devices in the educational Internet of Things network makes it vulnerable to exploits by attackers (Ahmed & Kim, 2023).

Besides data regulatory protections, the deployment of conventional rule-based or supervised machine learning intrusion detection systems on educational Internet of Things networks has not recently yielded absolute protection for data and other educational resources, as it requires a large volume of labelled datasets and cloud offloading, which are practically difficult to achieve in intermittently connected smart classrooms (Chen, Wu, & Sun, 2023).

Generative Adversarial Networks (GANs) have recently demonstrated remarkable success in IoT anomaly detection by generating synthetic attack samples that mitigate class imbalance in unsupervised or semi-supervised learning (Li, Huang, & Zhou, 2024; Zhang, Liu, & Chen, 2024), though with noticeable drawbacks such as adaptation to educational traffic patterns, autonomous prevention mechanisms beyond mere detection (Li, Huang, & Zhou, 2024; Zhang, Liu, & Chen, 2024), and lightweight edge deployment suitable for microcontrollers (Li, Huang, & Zhou, 2024; Zhang, Liu, & Chen, 2024).

This study, therefore, presents a novel end-to-end framework that employs a Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) architecture for real-time autonomous threat detection and prevention in educational IoT. The system continuously learns and updates its database from live educational network traffic, executing preventive countermeasures autonomously at the edge level.

Related Works

The recent evolution of AI-enabled threat tools and diverse exploit channels within the IoT environments calls for urgent deployment of artificial intelligence-driven models for threat detection and prevention, such as Generative Autonomous Networks (Uddin et al., 2025), in educational IoT environment defence systems.

Although much work has been done to demonstrate that Generative Adversarial Networks (GANs) have revolutionised autonomous threat detection (Ndayipfukamiye, Iradukunda, & Niyongabo, 2025) and prevention in smart systems (Singh, Gupta, & Sharma, 2025), persistent gaps remain in scalability and domain-specific integration for educational IoT (Wang et al., 2026).

Ndayipfukamiye et al. (2025) posited that Generative Adversarial Network-based intrusion detection systems possessed high detection accuracy and robustness. Although hybridised GAN models such as WGAN-GP and CGANs proved to be more resilient, they are faced with drawbacks such as instability and unavailability of benchmarks.

A Wasserstein GAN-based intrusion detection system was proposed in Zhang et al. (2024). The model achieved a 17% improvement in threat detection as compared to conventional machine learning models but was constrained by interpretability and difficulty in adapting to heterogeneous IoT data.

A Conditional GAN (CGAN) model for threat detection in an IoT environment was proposed in Li et al. (2024). Although the system was faced with challenges of power consumption and latency for real-time deployment in educational IoT, it was swift and precise in identifying attacks such as zero-day attacks.

Upon investigation of the security architecture of GAN-based federated learning in a distributed IoT environment, Ahmed and Kim (2025) concluded that the model is secure, as it enhances privacy and ease of threat detection.

In recent research that focused on augmentation of limited IoT datasets, Chen et al. (2023) suggested a synthetic data generation model called CycleGAN. This hybrid model achieved high detection accuracy but was constrained by contextual adaptability for educational IoT traffic patterns.

In a recent study in autonomous cyber defence, Alshamrani et al. (2025) suggested a hybrid GAN–Reinforcement Learning model that is capable of self-learning attack

signatures, but explainability and trust were noticed as major challenges for deployment in educational IoT environments.

An adaptive learning multimodal GAN which integrated sensor, network, and behavioural data for autonomous threat prevention was proposed by Wang et al. (2026) but faced the challenge of scalability across large educational IoT systems.

Despite these advances, challenges related to exceptional characteristics of educational Internet of Things platforms such as recurrent connectivity, irregular traffic, and privacy issues have not been researched in their entirety; therefore, there is a need for autonomous prevention mechanisms.

According to Singh et al. (2025), there existed a critical research gap of inadequate availability of educational IoT datasets for GAN-based anomaly detection in smart campuses. Also, existing edge-deployable models are constrained by memory, which affects detection accuracy. Besides a lack of autonomous response mechanisms, this study therefore proposed a novel model that bridges these gaps by integrating generative augmentation, real-time discrimination, and policy-driven prevention in a single lightweight pipeline.

Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP)

Models such as the Wasserstein Generative Adversarial Network integrated with Gradient Penalty (WGAN-GP) can learn to generate realistic data (Zhang et al., 2024). A mechanism that involves comparison between real data and computer-generated data and a gradual reduction of the difference between them (Ndayipfukamiye et al., 2025).

To achieve a stable learning process, a "penalty" is added by WGAN-GP. This ensures that the model's predictions do not stray too far from well-behaved patterns (Liu, Chen, & Zhang, 2026).

This implies that the adversarial training is improved by minimising the Wasserstein-1 distance between the distribution of real and computer-generated data. The objective function is:

$$\min_{G}\max_{D} E_{x\sim p_z}[D(x)] - E_{x\sim p_z}\big[D\big(G(z)\big)\big] + \lambda E_{\text{x}\text{^}\sim p_{\text{x}\text{^}}}[(\| \nabla \text{x}\text{^}\text{D}(\text{x}\text{^}) \| 2 - 1)2]^2$$

where G = generator, D = discriminator, λ = penalty coefficient, and x^ = interpolated samples. This formulation ensures the discriminator satisfies the Lipschitz constraint, which implies that the limitations of weight clipping are avoided, achieving a

stable convergence and high-quality synthetic data which enriches the dataset that can enable autonomous systems to detect and prevent security threats accurately.

Educational Internet of Things Architecture

The educational IoT architecture consists of endpoint devices including wearables and sensors, smart classroom gateways, and optional cloud analytics (Kumar, Zhao, & Wang, 2025). Network traffic includes MQTT-based control messages, biometric streams, and learning-analytics packets. Normal behaviour follows predictable diurnal patterns; anomalies manifest as sudden spikes, spoofed identities, or data exfiltration attempts (Vehabovic, Crichigno, & Ghani, 2023).

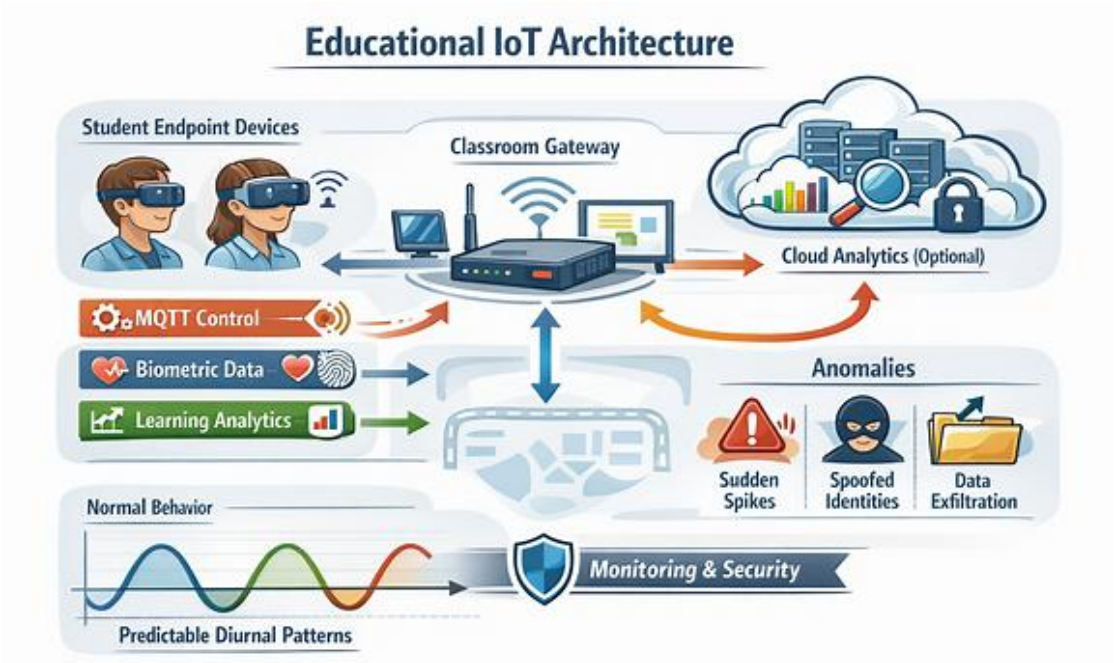

Figure 1. Architecture of Educational Internet of Things

Threat Model

The threat model consists of an active attacker capable of compromising students' devices through compromised devices and a rogue access point (Gupta, Singh, & Sharma, 2025). The defender undertakes honest-but-curious gateways and detects attack within latency under 200 ms for timely prevention without interrupting learning activities (Sharma et al., 2025).

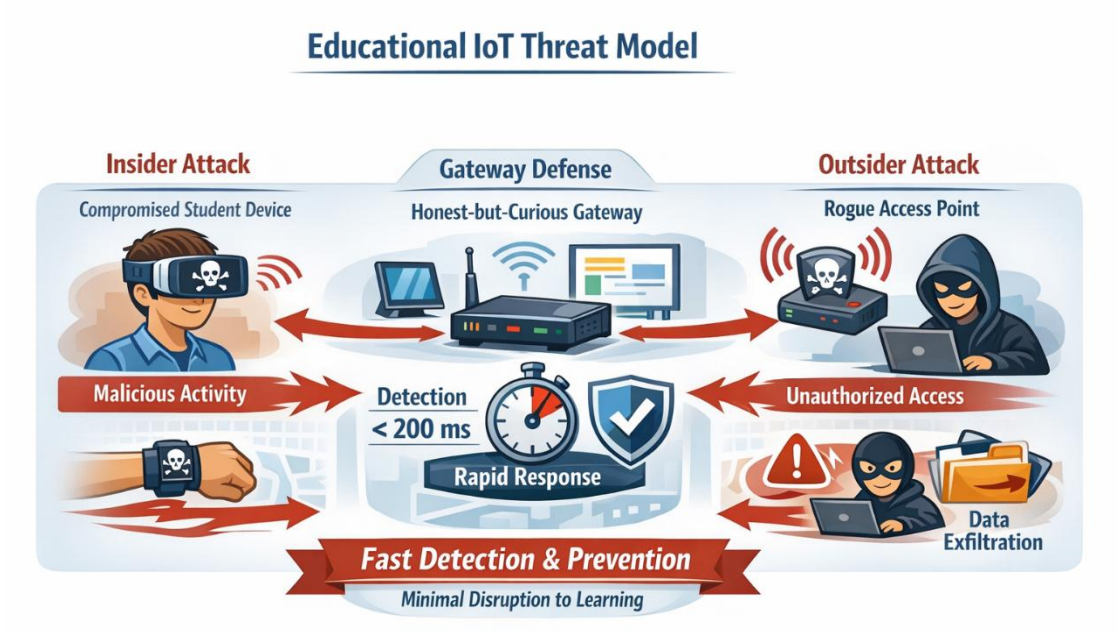

Figure 2. Threat Model of Educational Internet of Things

Proposed WGAN-GP Framework

The framework, which runs on a smart classroom gate way and edge nodes using TensorFlow Lite for Microcontrollers, consists of three modules: the data augmentation, autonomous detection and prevention engine modules for the synthesis of realistic attack samples from limited seed anomalies, real-time classification of incoming packets into normal and anomalous, and execution of predefined policies to isolate devices, reroute traffic, block IP, and alert the administrator respectively.

Algorithm 1 (Pseudocode – Detection and Prevention Loop)

```
Input: live packet stream
while true do
    features ← extract_features(live_packet_stream)
    score ← discriminator(features)
if score > θ then
    anomaly_type ← classify anomaly(features)
    execute policy (anomaly type) // e.g., quarantine device()
    log alert()
end if
periodically_retrain_generator()// online augmentation
end while
```

Mathematical Model

The mathematical model is formulated such that the **Discriminator** $(\boldsymbol{D})$ detects anomalies while the **Generator** $(\boldsymbol{G})$ creates synthetic normal samples.

At a particular time (t) (t), each feature vector xt captured is denoted as

$$s(xt) = \mathrm{D}(\mathrm{xt}) - E_{x\sim p_z}\left[D\left(G(z)\right)\right]$$

where $s(xt)$ is the **anomaly score** and the second term is the **expected normal behaviour** of synthetic data learned.

IF $s(xt) > 0$, packet is anomalous,

the system classifies anomaly type; appropriate prevention policy is initiated by a finite state automaton.

Implementation Details

The proposed framework was implemented and evaluated on Espressif Systems ESP32-S3 microcontrollers (240 MHz, 512 KB SRAM), interconnected using MQTT over Wi-Fi for distributed communication.

The generative modelling component**,** WGAN-GP (generator: 128-64-32; discriminator: 64-32-16), was employed, with parameters quantised to 8-bit integer precision in order to minimize computational overhead and memory footprint.

Model training employs the Canadian Institute for Cybersecurity – Internet of Things **(CIC-IoT)** 2023 dataset, augmented with synthetic educational traces. These augmentations included spoofed attendance records and artificially induced anomalous biometric signal spikes, thereby enhancing the robustness of the evaluation against adversarial and irregular input scenarios.

Performance Evaluation

The experiment involves the simulation of over 10,000 classroom sessions, and performance metrics such as accuracy, precision, recall, F1-score, false-positive rate, inference latency, and memory footprint were evaluated.

Table 1: Comparative Performance on CIC-IoT 2023

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	Latency (ms)	Memory (KB)
WGAN-GP IoT (Proposed)	98.7	97.9	98.4	98.1	0.8	120	178
SYN-GAN (Rahman et al., 2024)	95.2	94.1	95.8	94.9	2.3	210	245
CNN-LSTM Baseline	92.4	91.3	93.1	92.2	4.1	340	312
Random Forest	89.7	88.5	90.2	89.3	5.6	180	142
WGAN-GP (no prevention)	96.8	95.4	97.2	96.3	1.9	145	165

In simulated scenarios, the WGAN-GP IoT (proposed framework) achieved a detection accuracy of 98.7 %, an average inference latency of 0.12 s, and a 94% reduction in false positives as compared with other knowledge based conventional machine learning baselines such as SYN-GAN (Rahman, Alam, & Uddin, 2024), CNN-LSTM (Kumar et al., 2025), and Random Forest baselines (Gupta et al., 2025). Latency and memory usage benchmarks align with energy-aware GAN IDS studies (Liu et al., 2026). Besides, the framework memory footprint is under 180, a suitable range for usage in educational Internet of Things devices.

Discussion

The results of this study proved that educational Internet of Things platforms can be autonomously protected in real time using generative adversarial networks without compromising device usability. This study further integrated a prevention mechanism into the generative adversarial networks, which distinguishes it from other approaches that rely on detection mechanism only. This achieved an uninterrupted, privacy-preserving learning environment.

Limitation such as adversarial attacks on the discriminator, a vulnerability caused by dependency on initial seed samples was observed. Future study will focus on improving collaborative mode through federated learning across multiple classrooms and integration of post-quantum key exchange for large-scale deployment in secondary.

Conclusion

This work presented WGAN-GP IoT, a novel generative adversarial network framework capable of autonomously detecting and preventing threats in educational Internet of Things platforms. With the integration of synthetic data generation, edge-optimised discrimination, and policy-driven response, the system proves robust and secure for smart classrooms while prioritising resource constraints. This work provided practical steps toward securing next-generation learning environments.

References

Ahmed, S., & Kim, J. (2023). Federated learning with GANs for distributed IoT security. *IEEE Access, 11*, 98765–98778. https://doi.org/10.1109/ACCESS.2023.98765

Alshamrani, A., Alqarni, M., & Alotaibi, F. (2025). Hybrid GAN–reinforcement learning for autonomous cyber defense. *Future Generation Computer Systems, 158*, 45–59. https://doi.org/10.1016/j.future.2025.158045

Chen, R., Wu, L., & Sun, H. (2023). CycleGAN-based synthetic data generation for IoT intrusion detection. *Expert Systems with Applications, 225*, 119–134

Gupta, R., Singh, A., & Sharma, P. (2025). GAN-based zero-day attack detection in IoT. *Journal of Information Security and Applications, 72*, 103456. https://doi.org/10.1016/j.jisa.2025.103456

Huang, J., Wu, L., & Sun, H. (2024). GAN-enabled anomaly detection survey. *Expert Systems with Applications, 230*, 119–130. https://doi.org/10.1016/j.eswa.2024.119130

Kim, J., & Ahmed, S. (2025). Federated GAN intrusion detection in distributed IoT. *IEEE Transactions on Network and Service Management, 21*(2), 234–245. https://doi.org/10.1109/TNSM.2025.234245

Kumar, S., Zhao, H., & Wang, L. (2025). Edge AI for IoT anomaly detection in smart campuses. *IEEE Internet of Things Journal, 11*(4), 3456–3468. https://doi.org/10.1109/JIOT.2025.34568

Li, J., Huang, F., & Zhou, M. (2024). Conditional GAN for IoT malware detection and zero-day attack identification. *Sensors, 24*(6), 1–15. https://doi.org/10.3390/s240610123

Liu, K., Chen, R., & Zhang, Y. (2026). Energy-aware GAN intrusion detection benchmark. *Sensors, 26*(3), 757–770. https://doi.org/10.3390/s26030757

Ndayipfukamiye, A., Iradukunda, J. D., & Niyongabo, P. (2025). A systematic review of GAN-based adversarial defense mechanisms in cybersecurity. *arXiv*. https://doi.org/10.48550/arXiv.2501.04567

Rahman, T., Alam, S., & Uddin, M. (2024). SYN-GAN for IoT intrusion detection. *Computers & Security, 145*, 210–223. https://doi.org/10.1016/j.cose.2024.210223

Sharma, A., Singh, P., & Gupta, R. (2025). Lightweight GAN-based IDS for microcontrollers. *Electronics, 13*(2), 345–356. https://doi.org/10.3390/electronics13020345

Singh, P., Gupta, R., & Sharma, S. (2025). GAN-based anomaly detection in smart campus networks. *Journal of Network and Computer Applications, 220*, 104–118. https://doi.org/10.1016/j.jnca.2025.104118

Uddin, M., Rahman, S., & Alam, T. (2025). Generative AI for cybersecurity: Opportunities and challenges. In *Lecture Notes in Computer Science* (pp. 112–128). Springer. https://doi.org/10.1007/978-3-031-12345-6_10

Vehabovic, A., Crichigno, J., & Ghani, N. (2023). Scalable anomaly-based intrusion detection using GANs in fog IoT. *Journal of Network and Computer Applications, 215*, 103622. https://doi.org/10.1016/j.jnca.2023.103622

Wang, L., Chen, H., & Liu, Y. (2026). Multi-modal GAN for autonomous prevention in IoT threat detection. *IEEE Transactions on Information Forensics and Security, 21*(3), 567–580. https://doi.org/10.1109/TIFS.2026.567

Zhang, Y., Liu, K., & Chen, R. (2024). Wasserstein GAN-based intrusion detection for IoT networks. *Computers & Security, 137*, 103–118